

---

# Identification and Evaluation of Existing Nursing Homes Quality Indicators

Katherine Berg, Ph.D., Vincent Mor, Ph.D., John Morris, Ph.D., Katharine M. Murphy, Ph.D., R.N., Terry Moore, M.P.H., R.N., and Yael Harris, M.H.S.

---

*We summarize work done to identify and evaluate existing quality indicators (QIs) for long-term care (LTC) settings. Indicators operationally defined using routinely collected and computerized patient assessments were identified and then aggregated to characterize the performance of the nursing facility over a specific period of time. Of 143 indicators reviewed, only 22 were recommended for use in comparing performance across facilities. Conceptual and technical issues influence the appropriateness of QIs for different audiences.*

## INTRODUCTION

Facility-specific reports on quality of care can be essential new tools to enhance quality and public accountability of health care services. Such reports can be used to benchmark facility performance with peer facilities within a State or nationally. Intermediate purchasers of care such as Medicare, Medicaid, or managed care organizations might use such external quality reports to avoid contracting with poor providers.

Public reporting of facility performance, an important new development, is now being piloted. The intent is to spur providers to compete on the basis of quality

and to assist patients and their advocates in making informed decisions. Consumers, however, may not have expertise in interpreting QIs, and will not have the same level of access as surveyors to verify or refute impressions of facility performance gained from public reporting. What to report and how to present the information is, therefore, a matter of ongoing debate among experts. But no matter which decisions are taken to present the data, the development of QIs in multiple domains of quality is essential in meeting future public reporting goals.

To be useful for future public reporting and current regulatory and quality monitoring functions, QIs must meet certain measurement criteria including content validity, consistency over time periods, and validity in terms of representing quality within certain domains of care. Table 1 presents the list of desired measurement properties of QIs. The list may be used as a guideline for appraising and critically reviewing existing QIs. However, QIs may be useful for internal quality monitoring without demonstrating all the measurement properties listed in Table 1.

The audience for internal quality reports is intimately familiar with the facility, its patients, and care patterns. There is little risk of misinterpretation attached to quality reports as internal documents because readers will have ready access to the facility staff and patients to verify impressions left by the report. In that sense, the QI is a key starting point in the process of identifying and responding to opportunities to improve care.

---

Katherine Berg is with the McGill University School of Physical and Occupational Therapy and Brown University. Vincent Mor is with Brown University. John Morris and Katharine Murphy are with the Hebrew Rehabilitation Center for Aged. Terry Moore is with Abt Associates. Yael Harris is with the Centers for Medicare & Medicaid Services (CMS). The research in this article was supported by HCFA under Contract Number 500-95-0062. The views expressed in this article are those of the authors and do not necessarily reflect the views of McGill University, Brown University, the Hebrew Rehabilitation Center for Aged, Abt Associates, or CMS.

**Table 1**

**Desired Measurement Properties of a Quality Indicator (QI) for External Comparisons**

- Demonstrate consistency over adjacent time intervals when tested in multiple nursing facilities in multiple States.
- Have sufficient numbers of residents at risk (in the denominator) of the QI.
- Demonstrate sufficient prevalence or incidence when tested in nursing facilities.
- Address areas of quality of care that are important to consumers, health care professionals, nursing home administrators, long-term care researchers, and other quality-of-care experts.
- Have content validity in QI definition including numerator, denominator, clinical exclusions to denominator, and covariates used for risk adjustment.
- Have appropriate risk adjustment in the definition of the QI to adjust for potential biases related to referral/admission patterns, differential discharge or censoring rates, and resulting heterogeneity of patients left residing in the nursing facility.
- Not be unduly influenced by ascertainment bias arising from differences in assessment skills or vigilance of staff when assessing patients.
- Be under control of the facility, that is, the facility can improve the rate with improved practices in quality of care in that domain.
- Demonstrate responsiveness or ability to detect meaningful change in facility performance.
- Demonstrate relationships in expected direction with quality-of-care practices or other theoretically hypothesized relationships (construct validity).

SOURCES: Berg, K., McGill University School of Physical and Occupational Therapy and Brown University, Mor, V., Brown University, Morris, J., and Murphy, K., Hebrew Rehabilitation Center for Aged, Moore, T., Abt Associates, and Harris, Y., Centers for Medicare & Medicaid Services, 1998-2001.

QIs are used by facility surveyors in order to evaluate the processes and outcomes of care provided by a particular organization. In this case, the audience for the report is external. However, specially trained individuals external to the service organization typically conduct site visits during which they have the opportunity to interview staff, administrators, patients, and/or family members to gather additional information on the organization's care patterns and performance.

Before reviewing existing QIs, we propose a series of ratings on the degree to which various measurement properties are relevant for specific audiences: internal nursing facility staff, surveyors or external regulators, and for consumers. Table 2 presents the ratings for each type of audience. The ratings are consistent with the demands for more stringent measurement properties when QIs are used for public reporting than for surveyors or for internal quality monitoring.

This article addresses the early phases of work under a contract with CMS to develop and validate QIs for post-acute and LTC settings. The project was designed to assist CMS in advancing their vision for stimulating quality of care in nursing facilities by developing and validating QIs that

reflect clinical and other care outcomes at the facility level. Prior to developing new QIs, the first task of this project was to assess existing QIs and to determine which of them, if any, could be recommended to CMS for immediate use. While previous work has focused primarily on how QIs could augment the regulatory process, the goal of this project was to identify QIs for multiple audiences, ranging from nursing facilities themselves to the consumers and purchasers of care.

### **Measuring Quality in U.S. Nursing Homes**

One aspect of accountability of particular interest to CMS is patient outcomes. Care outcomes may now be accessed remotely via electronic database surveillance systems for select providers of service. This makes affordable and timely information on patient care outcomes available to CMS as both a purchaser of service and a regulator of quality care for Medicare beneficiaries. For example, Medicare+Choice organizations are monitored via the National Committee on Quality Assurance measures. Nursing facilities have recently begun to be monitored via QIs constructed from data elements from the Minimum

**Table 2**  
**Guide for Determining Acceptability of a Quality Indicator (QI) for a Particular Audience**

Focus of Measurement	Surveyor		
	Nursing Facility Internal QI Monitoring, Benchmarking	External QI – Drives	Public Reporting
		Survey Process, Accountability, Benchmarking	Communicating QIs to Consumers and Purchasers
Consistency of QI Over Time Intervals, e.g., 3-Month Periods Potential for Censoring Bias (Differential Discharge Rates, Hospitalization Rate)	0	+	+
Potential for Selection Bias (Differential Admission and Referral Practices)	0	++	++
Risk-Adjustment Adequacy	+	+	++
Face/Construct Validity of the QI Components	+	+	+
Reliability of Variables Scales Used in the QI	+	++	++
Degree of Potential Control by Facility Over the Outcome	+	+	0
Consistency of QI Over Multiple States	0	+	++
Importance and Relevance of the QI (i.e., the “So What” Test)	+	+	++

NOTE: 0 is little or no importance, + is important, and ++ is very important.

SOURCES: Berg, K., McGill University School of Physical and Occupational Therapy and Brown University, Mor, V., Brown University, Morris, J., and Murphy, K., Hebrew Rehabilitation Center for Aged, Moore, T., Abt Associates, and Harris, Y., Centers for Medicare & Medicaid Services, 1998-2001.

Data Set (MDS), a comprehensive assessment instrument used by all U.S. nursing facilities. Medicare-certified home health agencies are to be monitored for their care outcomes via outcome-based QIs, derived from the newly-mandated Outcome and Assessment Information Set assessment instrument.

In the past, researchers have used aggregate data about a facility as a basis for judging quality, but generally only for a small number of facilities or in select groups of facilities (Zinn, Aaronson, and Rosko, 1993a, b; Zinn, 1994; Nyman, 1988). The use of these data limits these studies, in that risk adjustment is quite minimal due to the ecological fallacy, i.e., inappropriate inferences drawn about individuals based on an aggregated database rather than patient-level information. This is one of the reasons that so many of the early studies of the determinants of quality of care in nursing facilities led to contradictory findings (Davis, 1991).

The introduction of the MDS into U.S. nursing facilities significantly altered the care quality arena. State-mandated computerization of MDS assessments in more than 10 States in the early 1990s helped to

create an MDS database from which QIs could be developed. The mandate in June 1998 for computer transmission of MDS assessments in all Medicare-certified nursing facilities across the Nation lead to the creation of State and national repositories for these MDS computerized data. With these resources, planners began to describe a system for designing, testing, and widely disseminating QI systems for nursing facility-based care.

The MDS is an assessment instrument that addresses multiple domains including cognitive function, sensory, physical function, pain, incontinence, skin problems and pressure ulcers, nutrition, diagnoses, signs and symptoms, special treatments, and medication use. The unique feature of the MDS is that it uses responses to various items to signal a potential problem, and suggests guidelines as to how that problem can be managed without being prescriptive. It is thus an interactive assessment instrument. Subscales within the MDS are also used to monitor status and as outcome measures.

The reliability of MDS assessments has been repeatedly demonstrated (Morris et al., 1990; Hawes et al., 1995; Morris et al.,

1997), and there is evidence supporting the validity of MDS items and subscales when compared with standardized research instruments that measure similar constructs (Morris et al., 1994; Hartmaier et al., 1994; Frederiksen, Tariot, and De Jonghe, 1996; Fries et al., 2001). Moreover, a recent series of studies using MDS data from five States participating in CMS's Nursing Home Casemix and Quality demonstration confirms the validity of the diagnostic and functional outcome data in the MDS (Gambassi et al., 1998; Bernabei and Gambassi, 1998; Bernabei et al., 1998; and Landi et al., 1998, 1999). The reliability and validity used to construct a QI are important, but not sufficient criteria for the reliability of QI performance.

The multi-dimensional nature of MDS assessments offers a unique opportunity to characterize the performance of U.S. nursing homes in multiple domains of care by aggregating patient-level MDS assessment data within a specific timeframe. For the LTC resident in a nursing facility setting, the aim is not simply to minimize the duration of treatment nor to expedite a return to some prior level of functioning. In fact, curative goals are often secondary to goals related to the prevention of decline. Facility performance should be considered in multiple domains such as its ability to prevent decline in physical and cognitive function, to manage clinical complexity, and to prevent adverse outcomes such as pressure ulcers, and to foster psychosocial well-being. In this regard, the nursing home sector has an enormous advantage over the ambulatory and the acute-care sectors since nursing facilities now have electronically collected MDS data with clinically meaningful resident-level measures on an ongoing basis for all residents.

Work in the area of nursing facility quality has also explored the use of information beyond the MDS. Harrington and col-

leagues (1999a, b) examined the relationship between Online Survey Certification and Reporting System data and staffing standards in nursing facilities. Kramer and associates (1999) have developed 80 QIs that use a combination of MDS, nursing records, staff interviews, and resident observations. Development of quality-of-life indicators for nursing home residents is being lead by Kane and colleagues (2000) from the University of Minnesota. Development and validation of QIs in these areas were not within the mandate of the present study.

In this article, we report on the review of existing QIs and, the preliminary analyses that examined the incidence and prevalence of selected QIs in nursing facilities across multiple States, and the stability of the QI rates over time. The early phases reported here were not mandated to make overall determinations of facility quality or address all desired measurement properties listed in Table 1. Future phases of the project are planned to develop QIs in areas not addressed by existing QIs and to examine validation strategies with onsite data collection.

The methods and results of the study are organized into two separate sections. The first describes the methods and results used to review the literature and determine which indicators warrant further analysis. The second section describes the methods and results of the preliminary analyses using National Repository MDS data.

## **Review of Existing QIs**

The objective of this phase was to identify all possible QIs appropriate for use in nursing facilities and to determine to what degree the existing QIs meet the essential measurement criteria when used with chronic patients in nursing facilities. Indicators based on MDS items or ones

that could be formulated using MDS items were preferred because the data were readily available for all U.S. nursing homes in the computerized CMS National Repository.

## Methods

Sources of information included the published research literature, listings of ORYX™ vendors, Web sites for nursing home quality, and LTC quality researchers. Many ORYX™ vendors who contract with nursing facilities use MDS-based QIs to monitor facility performance. Monitoring by approved ORYX™ vendors is a requirement of the Joint Commission on the Accreditation of Healthcare Organizations for nursing facilities that wish to be accredited.

The first step in the review process was the construction of review teams for each of the substantive domains into which the QIs had been classified (function, clinical complexity, psychosocial, pharmacotherapy). Each team consisted of clinicians with LTC content expertise in the domain as well as experienced researchers aware of the technical issues associated with constructing QIs.

In total, 57 QIs from published articles and 112 from ORYX™ vendors were identified and forwarded for review. Each reviewer received all articles and information pertinent to a specific area such as pressure ulcers, incontinence, or functional decline. Each different QI was entered as one record into a computerized tracking system—an ORACLE database. Reviewers made the determination of which QIs represented unique markers. In total, 143 QIs were entered into the database. However, later clarifications from organizations revealed certain records were duplicates. That is, other organizations submitted QIs that were identical to QIs at the Center for Health Service Research and Analysis

(CHSRA) at the University of Wisconsin.

Reviewers also had to specify the technical aspects of the QI including the numerator, denominator, and risk-adjustment method. The numerator refers to the upper portion of a fraction used to calculate a rate, proportion, or ratio, i.e., patients who have the characteristic or outcome of interest. The denominator refers to the lower part of a fraction used to calculate the rate, proportion, or ratio. It may refer to a subgroup at risk of developing the outcome or characteristic of interest or may refer to all persons at risk within a facility. Risk adjustment may consist of one or more of three basic types: (1) restricted denominator, (2) separate calculation of QI rates within risk groups or strata, and (3) use of multivariate adjustment modeling. Reviewers were also asked to document available information on the measurement properties of the QI, including reliability, validity, and responsiveness.

In cases with incomplete or unclear indicator definitions, study personnel contacted developers via letters and with followup telephone calls.

## Results of Review

Relatively little information on measurement properties was available from the published literature. Few developers provided detailed information on the methods or results of assessing the performance of their QIs, i.e., the validation of their QIs. The majority of ORYX™ vendors provided only the definition of their QI(s), often without detailed instructions.

A notable exception was the work of Zimmerman and colleagues. They conducted both pilot testing and primary validation for the QIs they had developed. Specifically, they checked the accuracy of the data underlying the QI (Zimmerman, Karon, and Swearingen, 1999; Zimmerman

and Karon, 1997; and Zimmerman et al., 1995) and made a determination as to whether individuals who were identified as having a problem based on the QI actually had the problem and that the problem was related to actual quality-of-care problems at the resident or facility level (Zimmerman et al., 1999; Zimmerman and Karon, 1997). They also compared facility rankings within three States with the presence of any deficiency citations by State surveys conducted at the same time (Zimmerman et al., 1999). Results demonstrated that QIs with high rates of occurrence, selected at high threshold levels, are useful tools for identifying quality of care problems at both the facility and resident levels. The QIs generally had high accuracy ratings, and most identified severe problems for all or some residents.

The CHSRA team identified the need for further research in identifying possible QI-specific thresholds. In addition, the researchers noted a need for further studies to examine observed differences between specific deficiencies assigned by the validation teams and those assigned by State surveyors onsite at the same time. Thus, although QIs developed by CHSRA are the most widely used and are the ones for which there is the greatest amount of information in the published literature, documentation is still lacking relative to most of the measurement properties listed in Table 1.

LTCQ, another ORYX™ vendor, also provided preliminary evidence of validity for multiple QIs. They reported a relationship between QIs pertaining to rates of decline with survey deficiency rankings by State. They also observed a relationship between selected facility characteristics (e.g., structural features such as levels of staffing), and other QIs obtained from the annual State surveyors' inspections. Lastly, LTCQ reported that facilities with

better rates of QIs had higher global ratings of innovative and best practices when judged by peer facilities.

In summary, although there was little or no documentation for many QIs currently in use, the fact that they were used suggested some degree of content validity and perceived clinical utility. In addition, certain developers demonstrated accuracy and presented evidence supporting the validity of their indicators. The results underscored the need for further testing of the QIs, but certainly supported the direction of this research.

### **Selection for Empirical Testing**

Based on the results of the review process, the project steering committee determined that the minimum criteria for selection would be the presence of a clearly specified numerator and denominator, both of which could be operationally defined using MDS items. An exact definition of a QI requires: precise instructions regarding which data elements constitute the measure, the methodology for aggregating the data, the data element combination of each composite measure, and any further instructions regarding stratification or risk adjustment. Without clear instructions and a rationale for these choices, it is difficult to know whether the QI proposed by the developers is being correctly replicated.

A priori the steering committee decided to give preference to QIs with some form of risk adjustment in order to permit a fairer comparison between facilities with different patient populations (or case mix). However, no QIs failed to be forwarded for empirical analysis solely on the basis of an absence of risk adjustment. Consideration was given to the perceived clinical relevance, presence of literature either supporting or opposing the concept and its

relationship to quality, whether the expected prevalence or incidence would be sufficient to function as a reliable QI over time and whether the indicator was likely to be under the control of the facility.

A total of 44 QIs were submitted for empirical analyses; most were derived either from CHSRA or from LTCQ's Q-Metrics®, Information Advisory System. Both QI systems have been used by facilities in multiple States for many years and are certified as ORYX™ vendors under the Joint Commission on the Accreditation of Healthcare Organizations outcomes measurement initiative. Moreover, 24 CHSRA QIs are incorporated into CMS's survey process, and, thus, were candidates for further testing on that basis.

### **Methods Used in Initial Empirical Analyses**

MDS assessments from 1995-1997 from all nursing facilities in Kansas, Maine, New York, South Dakota, and Vermont were used in the analysis. Full MDS assessments must be completed by day 14 for all patients admitted to nursing homes and thereafter annually. In addition, quarterly assessments based on a smaller subset of items are to be done every 3 months. Medicare prospective payment for nursing homes requires more frequent assessments for reimbursement, but no post-acute QIs were included in the present analyses.

#### *Resident File Construction*

These resident-level files contain four quarters of the calendar year 1996 data. For residents with multiple MDS assessments during the quarter, the priority was given to the full assessment, but if none was completed within the quarter, the most recent quarterly was selected for the file.

### **Facility Aggregate Counterparts to Resident Files**

These facility-level files contain aggregated measures, based on the residents in the quarterly and annual files. A minimum of 20 aggregated cases for the QI under review were required in order for the facility to be included in the analysis of that QI.

### **Rationale for Choosing Sample Size**

Creating a QI for a given facility requires aggregating data about residents, events, or treatment processes to the level of the facility. Although all appropriate observations may be included in the aggregate measure being constructed, a sample still determines the measure, since observations may change over time for any number of different reasons. In the case of constructing aggregated QIs, the best understanding of the cause of sampling error is the number of observations determining the indicator. The larger the denominator determining the estimate, the more likely that the observed score is reasonably close to the true score. Table 3 summarizes this relationship for a hypothetical QI that has a prevalence (or incidence) of only 5 percent. Confidence intervals for proportions of 0.25 and 0.45 were of similar width. For a QI (e.g., cognitive impairment), with a prevalence of 0.45, the confidence intervals surrounding an estimate based on only 20 residents are 0.23 - 0.68. For this reason, we specified that facilities needed a minimum of legitimate observations for a specific QI to be included in the empirical validation analyses.

The rationale for conducting empirical analyses of the selected QIs in multiple States relates to the fact that experience working with these data has shown that there are substantial interstate differences in the prevalence of certain clinical conditions

**Table 3****Relationship Between Sample Size and the Standard Error of Estimate for a Hypothetical Quality Indicator with Incidence of 0.05**

Number of Observations	Standard Error	95 Percent Confidence Intervals <sup>1</sup>
10	0.09	0.002 - 0.444
20	0.05	0.001 - 0.245
30	0.04	0.008 - 0.223
50	0.03	0.01 - 0.160
100	0.02	0.01 - 0.110
200	0.01	0.02 - 0.090
500	0.009	0.03 - 0.070

<sup>1</sup> Binomial exact.

SOURCES: Berg, K., McGill University School of Physical and Occupational Therapy and Brown University, Mor, V., Brown University, Morris, J., and Murphy, K., Hebrew Rehabilitation Center for Aged, Moore, T., Abt Associates, and Harris, Y., Centers for Medicare & Medicaid Services, 1998-2001.

or outcomes such as functional change rates. This had alerted the team to the real possibility that data from different States might yield different answers to the question of the adequacy of the QI. While experience with these data suggested that some of the observed interstate differences were attributable to real differences in the populations of nursing home residents from State to State, the project team recognized that there were also substantial interstate differences in the measurement and assessment approaches used by nursing facility staff. Thus, we expected to observe both inter- and intrastate differences in the approach to measurement of clinically relevant phenomenon affecting QI performance.

For each QI tested, an analysis report summarizing the prevalence of the QI and its distribution within and across at least three States' population-based MDS data was generated (data not shown). Facility QI rates were examined within States; specifically, the mean rate and standard deviation, as well as the rates at the 10th, 50th, and 90th percentiles within each State. QIs with mean rates below 2 percent were flagged for further discussion as to whether they really represented sentinel events rather than a QI. The definition of a sentinel event was—a rare but critical phenomenon that might signify immediate danger of harm for residents.

Descriptive analyses examined the frequency distribution of proposed covariates and potential confounders. Bivariate analyses examined the correlation between the demographic variables and proposed clinical covariates with the unadjusted QI. Collinearity among covariates for each QI was also checked. Multivariate modeling at the person-level examined the relationship between raw QIs (numerator and denominator without covariate adjustment) and potential covariates. Odds ratios of proposed covariates were examined to determine if they were in the hypothesized direction and if the relationship remained consistent across multiple States. All covariates proposed by developers were tested in the multivariate models. To be acceptable, a covariate had to show consistency in at least two State data sets, that is, the effect of the covariate on the QI had to be of approximately the same magnitude and direction across multiple States before we believed that the underlying relationship was generalized.

All facility-level quality rates were examined for stability over time (across quarters) in two ways: (1) by correlating rank order of the deciles of the two sets of QI scores; and (2) by comparing movement on terciles of the QI distribution (e.g., whether facilities in the top one-third of the State distribution in one quarter retained that designation in the next quarter).

## Methods of Risk Adjustment

The type of risk adjustment was consistent with the method suggested by the developers. Thus, if the developer had no risk-adjustment methods, none was used. If the developers used stratified risk adjustment, then descriptive statistics and consistency over time intervals was examined within each strata separately. Multivariate modeling was done to risk adjust the regression-based QIs.

To calculate QI rates adjusted for resident-level risk factors, the project team used the method described by Berlowitz et al. (1996). Using logistic regression, each resident's logarithmic odds of experiencing an event is modeled as a linear function of his or her clinical characteristics. After arithmetic transformation, a predicted event probability for each resident is calculated from this estimation. Those predicted probabilities are summed for each facility and then divided by the number of residents at risk for the respective event to retrieve an expected event rate for the facility. The ratio of the actually observed event rate and the expected event rate is multiplied by the grand mean event rate, i.e., the event rate across all facilities, to give the risk-adjusted QI rate. The corresponding formula is:

$$\text{QI adj} = (\text{QI obs} / \text{QI pred}) \times \text{grand mean}$$

Adjusted event rates based on this technique have the following useful properties:

- The better the facility is doing compared with the model's prediction, the better (lower) is its QI rate.
- The worse the facility is doing compared to the model's prediction, the worse (higher) is its QI rate.
- If the facility's observed QI rate is equal to 0, its adjusted QI rate is also equal to 0.
- The average adjusted QI rate is close to the average observed QI rate.

## Results of Empirical Analyses

Following the statistical analyses, the project steering committee discussed and thoroughly reviewed each QI in terms of its distributional characteristics, stability, and cross-State consistency. The balancing perspective adopted in reviewing QIs was not to look for QIs that met all established performance criteria since none were really without problems. Rather, the project steering committee considered those QIs that minimized the problems of ascertainment bias, censoring through differential discharge rates, skewed distribution, case-mix adjustment, and QIs that resulted in dropping too many facilities due to insufficient sample size. On the other hand, in certain instances the concepts identified in QIs that may not have met these standards were deemed to be so important from a clinical or operational perspective that clinical review teams recommended their use and the steering committee agreed to include them in the list of recommended QIs. The results for 26 indicators initially recommended for use are presented in Table 4 within four broad domains: (1) functional, (2) clinical complexity, (3) psychosocial, and (4) pharmacotherapy.

Table 5 presents the prevalence/incidence unadjusted and adjusted rates, the rates at 10th and 90th percentiles and estimates of the stability over time intervals of selected QIs from the four domains. In general, there is variation within and between States in mean QI rates as well as the rates at the 10th and 90th percentiles. The QI with the lowest rate was the prevalence of feeding tubes at 3 percent in Kansas and 8 percent in New York. The highest rate was for prevalence of bladder and bowel incontinence with 45 percent in Kansas and 65 percent in Maine.

**Table 4**  
**Summary of Quality Indicators (QIs) that Underwent Empirical Testing**

Quality Indicator Domain	Cross-Sectional QIs		Change in Status QIs	
	Total Reviewed	Number Initially Recommended	Total Reviewed	Number Initially Recommended
Functional Status QIs Total	1	0	9	4
<b>Communication/Cognition</b>				
Communication	0	0	1	1
Cognition	0	0	2	1
<b>Activities of Daily Living Status</b>				
Bedfast	1	0	0	0
Locomotion	0	0	1	1
Activities of Daily Living	0	0	3	1
Range of Motion	0	0	2	0
Clinical Complexity QIs Total	13	6	10	7
<b>Continence Related</b>				
Bladder/Bowel	2	1	2	2
Fecal Impaction	1	0	0	0
Catheter	1	1	1	1
Urinary Tract Infection	1	1	0	0
<b>Nutrition/Hydration</b>				
Dehydrated	1	0	0	0
Weight Loss	1	0	1	1
Tube Feeding	2	1	0	0
Restraints	2	1	1	0
<b>Falls/Fracture</b>				
Falls	0	0	2	1
New Fractures	0	0	1	0
Pain	0	0	1	1
Pressure Ulcers	2	1	1	1
Psychosocial QIs Total	4	4	3	3
Mood	2	2	1	1
Behavior	1	1	1	1
Activity	1	1	0	0
Personal Relationships	0	0	1	1
Pharmacotherapy QIs Total	4	2	0	0
Anti-Anxiety/Hypnotics	2	1	0	0
Anti-Psychotic	1	1	0	0
9 or More Medications	1	0	0	0

SOURCES: Berg, K., McGill University School of Physical and Occupational Therapy and Brown University, Mor, V., Brown University, Morris, J., and Murphy, K., Hebrew Rehabilitation Center for Aged, Moore, T., Abt Associates, and Harris, Y., Centers for Medicare & Medicaid Services, 1998-2001.

Consistency across time intervals was assessed by the Spearman rank order correlation between decile rankings of facili-

ties within State over successive quarters. Table 5 shows some degree of variation in the consistency of QI rates across time

**Table 5**  
**Distribution of Rates and Estimates of Stability Over Time Intervals for Selected Functional, Clinical Complexity and Psychosocial Quality Indicators<sup>1</sup>**

Quality Indicator	State	Raw Rate Mean	Adjusted Rate Mean	***Rate At Percentiles			Stability	Risk-Adjusted Method
				10th	90th	90th		
Activity of Daily Living Decline	Vermont	0.15	—	0.06	0.29	*0.01	None	
	Kansas	0.12	—	0.04	0.21	*0.35		
	New York	0.09	—	0.04	0.15	*0.41		
Decline in Locomotion	Maine	0.23	0.23	0.09	0.42	**0.38	Regression-Based Adjustment	
	Kansas	0.12	0.12	0.03	0.22	**0.49		
	New York	0.11	0.11	0.04	0.19	**0.47		
Prevalence of Bladder and Bowel Incontinence	Vermont	0.60	0.93	0.85	1.00	0.55	High/Low-Risk Groups	
			0.45	0.31	0.60	0.81		
	Maine	0.65	0.91	0.80	1.00	0.73		
			0.47	0.30	0.64	0.73		
	Kansas	0.45	0.86	0.70	1.00	0.78		
			0.31	0.14	0.47	0.85		
	New York	0.60	0.95	0.89	1.00	0.74		
			0.47	0.29	0.66	0.90		
	South Dakota	0.46	0.88	0.71	1.00	0.78		
			0.34	0.21	0.48	0.82		
Bladder Incontinence Incidence or Worsening	Vermont	0.20	0.20	0.08	0.30	0.43	Regression-Based Adjustment	
	Maine	0.20	0.20	0.08	0.31	0.30		
	Kansas	0.14	0.14	0.05	0.24	0.34		
	New York	0.14	0.14	0.07	0.22	0.33		
	South Dakota	0.15	0.15	0.06	0.25	0.34		
Pressure Ulcers	Vermont	0.05	0.06	0.00	0.14	0.96	Regression-Based Adjustment	
Incidence or Worsening	Kansas	0.04	0.05	0.00	0.10	0.97		
	New York	0.05	0.06	0.02	0.11	0.94		
Weight Loss Incidence	Vermont	0.08	0.08	0.02	0.15	0.21	Regression-Based Adjustment	
	Kansas	0.09	0.08	0.02	0.15	0.26		
	New York	0.07	0.07	0.00	0.14	0.20		

See footnotes at end of table.

**Table 5—Continued**  
**Distribution of Rates and Estimates of Stability Over Time Intervals for Selected Functional, Clinical Complexity and Psychosocial Quality Indicators<sup>1</sup>**

Quality Indicator	State	Raw Rate Mean	Adjusted Rate Mean	***Rate At Percentiles			Stability	Risk-Adjusted Method
				10th	90th	90th		
Prevalence of Daily Physical Restraints	Vermont	0.08	—	0.00	0.21	0.87	None	
	Kansas	0.05	—	0.00	0.13	0.83		
	New York	0.09	—	0.00	0.23	0.94		
Prevalence of Feeding Tubes	Vermont	0.03	0.04	0.00	0.10	0.65	Regression-Based	
	Kansas	0.02	0.03	0.00	0.08	0.59	Adjustment	
	New York	0.08	0.08	0.00	0.17	0.73		
Mood/Depressive Symptoms with No Treatment	Vermont	0.09	—	0.02	0.19	0.80	None	
	Maine	0.05	—	0.00	0.10	0.73		
	Kansas	0.04	—	0.00	0.12	0.75		
	New York	0.02	—	0.00	0.06	0.65		
	South Dakota	0.02	—	0.00	0.04	0.63		
Deterioration in Mood	Vermont	0.24	0.25	0.12	0.38	0.17	Regression-Based	
	Maine	0.17	0.17	0.07	0.29	0.56	Adjustment	
	Kansas	0.16	0.17	0.06	0.29	0.35		
	South Dakota	0.17	0.17	0.08	0.27	0.51		
	New York	0.10	0.10	0.03	0.17	0.57		

\*\*\*Values represent adjusted rates. If rates were not adjusted, raw rates were used.

\*\*Percentage of facilities changing 3+ deciles.

\*Range of intertime decile correlation.

<sup>1</sup> Examples from list forwarded for use to the Centers for Medicare & Medicaid Services.

<sup>2</sup> High risk.

<sup>3</sup> Low risk.

NOTE: For stability, unless noted, correlation is across two adjacent 3-month time periods.

SOURCES: Berg, K., McGill University School of Physical and Occupational Therapy and Brown University, Mor, V., Brown University, Morris, J., and Murphy, K., Hebrew Rehabilitation Center for Aged, Moore, T., Abt Associates, and Harris, Y., Centers for Medicare & Medicaid Services, 1998-2001.

periods. In general, rates of prevalence-based QIs (e.g., physical restraint use [correlation: 0.83-0.94]) were more consistent over time intervals than change-based indicators such as decline in activities of daily living (ADLs) (correlation over time: 0-0.41). In addition, high-risk strata demonstrated generally lower reliability across time intervals than the low-risk strata. The lack of stability over time periods likely related to the smaller numbers of individuals in the high-risk strata.

At this stage, 26 QIs were supported for use by CMS. The specific QIs are reviewed here grouped by domain: functional, clinical complexity, psychosocial, and pharmacotherapy.

### **Functional QIs Recommended for Use**

Four functional QIs were recommended for use. Each QI represented a decline in status from one assessment to the next quarterly assessment and excluded those who could not decline further because they had maximal scores at the start of the time period. Functional decline was measured by a two-level decline in eating, bed mobility, transfer, and toileting or a one-level decline in two or more of these “late-loss” ADLs. Decline in cognitive function was measured by any deterioration in the cognitive performance scale over time (Morris et al., 1994). This indicator was preferred to the CHSRA decline in cognition indicator because it was more broadly applicable. The CHSRA denominator only applied to individuals who were cognitively intact on admission, thus excluding approximately 60-80 percent of nursing home residents who have some degree of cognitive impairment on admission to nursing facilities. Deterioration in communication was

based on a decline in either one of two MDS communication items: making self-understood or ability to understand others. Locomotion was measured by combining wheelchair mobility and walking into a single variable, and examining residents with some degree of independence who became more dependent within the next 90 days.

Analyses were conducted separately within the State, with each QI tested on at least two States and most often on two or three States. The four recommended functional QIs show relatively low associations between scores over time, suggesting instability in rankings from quarter to quarter. Indeed, the ADLs decline indicator had the lowest correlations between intertime deciles (0-0.41).

None of the four QIs had a low rate of occurrence, as all were 0.08 or above. Mean adjusted and unadjusted rates for the QIs were very similar. There was marked variation at the 90th percentile rate across States for decline in communication, cognition, and locomotion, with one State having double the rate at the 90th percentile than the other States tested. ADLs decline had no risk adjustment beyond restricting the resident pool to residents with some degree of independent function at the baseline assessment. Cognition, communication and locomotion had regression-based risk adjustment models.

### **Clinical Complexity QIs Recommended for Use**

The steering committee initially voted to accept a total of 13 QIs (six cross-sectional; seven change in status) for recommended use by CMS. The recommended clinical QIs can be conceptually characterized as either representing resident symptoms/clinical conditions or clinical processes of care.

## QIs Related to Symptoms or Conditions

Nine recommended QIs are related specifically to resident symptoms or clinical conditions (i.e., three for incontinence, one for urinary tract infection, two for pressure ulcer, one for falls, one for pain, and one for weight loss.). Three measures of incontinence were recommended, one cross-sectional and two change of status measures. Prevalence of bowel and bladder incontinence QI examines the proportion of patients with frequent or greater incontinence of either type at a point in time. It is simple to follow, has face validity, and is able to distinguish residents at high or low risk for the condition. Bladder incontinence incidence or worsening QI measures a deterioration over time of bladder incontinence, a common sign of potentially reversible or treatable conditions in the nursing facility population (e.g., delirium, urinary tract infection, joint pain limiting self-toileting ability). Bowel incontinence incidence or worsening measures a deterioration in bowel incontinence, a potentially treatable problem that may be associated with underlying constipation, fecal impaction, or laxative use.

New or worsening incontinence is relatively common (14 to 20 percent in nursing homes) across States. While some decline in bladder or bowel continence may not be reversible or manageable in the latter stages of disease (e.g., dementia, terminal illness) these QIs as operationalized appear to have the capacity to identify facilities where there may be a quality problem.

Urinary tract infection is a common problem among frail, debilitated nursing facility residents, often leading to hospitalization and poor quality outcomes (e.g., delirium, incontinence, falls). The QI was recommended for further validation based on clinical importance, but the recommendation was made with caution as several

potential problems were noted including a risk of underreporting and potential for censoring bias due to discharge to hospital for treatment. This QI also lacks any risk adjustment.

Two measures for pressure ulcer were recommended, one cross-sectional and the other a change in status measure. The prevalence QI utilizes a high risk/low risk-adjustment procedure to identify residents with any stage pressure ulcer on the most recent assessment. As it excludes new admissions and readmissions, it focuses attention on the prevalence of ulcers occurring in the facility. The change-based QI measures incidence or worsening of a pressure ulcer.

QIs for new or worsening pain, falls, and weight loss are all QIs that represent common symptoms of potentially treatable or manageable underlying conditions, or in the case of pain and weight loss, conditions that would prompt palliative measures towards the end of life. Although there were some concerns with the specification of the pain and weight loss QIs, the steering committee determined that, because they were both serious quality issues in nursing facilities, and they were the only covariate-adjusted QIs representing these concepts, they should be recommended for further validation and could be used by CMS at this time because they serve to identify facilities that have problems in these areas.

## QIs Related to Processes of Care

Four of the recommended QIs represented clinical care processes: physical restraints, feeding tubes, prevalence of indwelling catheter, and insertion of an indwelling catheter. The indicator prevalence of physical restraints measures the proportion of residents who have daily use of limb or trunk restraint or sit in chairs

that prevents rising. There is variation in the facility rates of daily restraint use across States and the rate at the 90th percentile varied from 0.13-0.23 across States. Although this QI is not risk adjusted, it does identify facilities with higher than average use.

The feeding tube QI measures the prevalence of feeding tube use. Even with risk adjustment, there is wide variation in the overall rates of feeding tube use. Two measures of indwelling catheter were analyzed. The prevalence of catheter use is an unadjusted QI whereas the incidence of new catheters since the prior assessment is based on a covariate-adjusted model. Both yield variation in distribution rates across facilities. Because indwelling catheters are associated with iatrogenesis and morbid outcomes in this population and these are important QIs over which facilities have some control, the steering committee recommended both QIs.

### **Psychosocial QIs Recommended for Use**

Seven psychosocial QIs were recommended for use, four prevalence and three longitudinal QIs: prevalence of behavior, decline in behavior, prevalence of mood/depression symptoms, deterioration in mood, mood/depressive symptoms with no treatment, little or no activity, and unsettled personal relationships.

The behavior prevalence QI utilizes a high-risk/low risk-adjustment procedure, while the decline in behavior status QI uses a covariate model. The prevalence of mood/depressive symptoms QI and the mood with no treatment QIs are not risk adjusted. Ascertainment bias may partly explain the wide distribution of the facility rates for mood with no treatment within and across States. For examples, the 10th

to 90th percentiles were 0.02 to 0.19 in Vermont, and 0.0 to 0.04 in New York (Table 5).

Unsettled personal relationships QI is based on the annual MDS assessment. This QI uses a covariate model, but the steering committee recommended that future work on this QI focus on refinement of the adjustment model. The little or no activity cross-sectional QI utilizes a restricted resident pool as risk-adjustment method.

### **Pharmacotherapy QIs Recommended for Use**

Two of the four pharmacotherapy QIs developed by CHSRA were recommended for the set of QIs to be provisionally recommended for use by CMS. Both recommended measures are cross-sectional and one of them is case-mix adjusted. The two QIs recommended for further use pertain to antipsychotic drug use and to the use of anti-anxiety/hypnotic agents. Antipsychotic drugs have been called pharmacological restraints since they are administered to residents with behavioral problems, hallucinations, and verbal outbursts associated with brain diseases such as Alzheimer's. The fact that the prevalence of antipsychotic-use QI excludes residents with selected psychiatric diagnoses from consideration and then differentiates between high- and low-risk residents addresses a number of problems that arise because some facilities have historically admitted residents with psychiatric histories. The unadjusted rate for antipsychotic use was approximately 14 percent in all States, but there was marked variation at the 90th percentile within strata.

The anti-anxiety/hypnotic drug use QI is not stratified, but does exclude residents with selected psychiatric diagnoses. The problematic aspects of anti-anxiety use as a

global indicator of quality is that it is not sufficiently precise. While many such drugs can be inappropriate for older persons, the greatest potential for damage lies in receipt of long acting, high dose benzodiazapines. These have been associated with falls and hospitalization for hip fracture, but, depending on the facility, may only represent a minority of all anti-anxiety/hypnotic use. Nonetheless, the QI as currently operationalized appears to provide the basis for identifying facilities that may have high use of this more restricted and problematic class of drugs, something that can be checked more completely through onsite inspection.

## **SUMMARY AND CONCLUSIONS**

There are numerous conceptual and statistical issues associated with the measurement of quality and the development of QIs. Interfacility comparisons must consider case-mix differences to address the selection bias that threatens the validity of the comparisons. Case-mix differences can occur due to differential admission practices and to differential rates of discharge (censoring due to death, transfer, or discharge home). Facilities may also be wrongly classified as performing better or worse due to assessment measurement errors. Such misclassification may result from facility differences in assessment skills, whereby better facilities detect more residents with a given problem and thus, appear to be performing more poorly than facilities that have low rates because they failed to assess the problem. Facilities may also be misclassified if the QI rates used to assess their performance are unstable due to small numbers of residents included in the calculation of the QI. For example, facilities that are ranked within 30 percentile scores of each other may only have a difference of absolute rates of 0.05. To

date, to our knowledge, no set of QIs in the acute, ambulatory, or the LTC arena has been adequately tested to make sure that it fully addresses all of these issues.

Conceptual considerations and technical analyses began during the initial phase of the project, continued in preparation for the next phase. Although not formally presented in this article, the additional technical analyses helped inform the final recommendations of which QIs were ready for use and appropriate for further validation. As a result, four QIs initially supported were not recommended for use to CMS. The four QIs include prevalence of mood/depressive symptoms and mood/depressive symptoms without treatment, unsettled personal relationships and prevalence of anti-anxiety/hypnotic use.

We do recognize that the 22 recommended QIs do not overcome all the methodological problems enumerated previously. However, the project steering committee believed that use of the QIs in the survey process and in some cases in public reporting adds considerably to regulators' and consumers' knowledge base. In spite of obvious limitations, it is the project team's opinion that using these adjusted QIs is better than not using them since they propel the industry forward while not unduly penalizing facilities for real differences in admission as well as assessment practices.

The multi-dimensional nature of quality should also be kept in mind when reporting QIs to any audience. Further work is underway for defining new QIs and testing all proposed indicators in an extensive onsite validation project.

## **ACKNOWLEDGMENT**

The authors would like to thank Sue Nonemaker, who was our project officer at the outset.

## REFERENCES

- Berlowitz, D.R., Ash, A.S., Brandeis, G.H., et al.: Rating Long-Term Care Facilities on Pressure Ulcer Development: Importance of Case-Mix Adjustment. *Annals of Internal Medicine* 124(6):557-563, March 15, 1996.
- Bernabei, R., and Gambassi, G.: The SAGE Database: Introducing Functional Outcomes in Geriatric Pharmaco-Epidemiology. *Journal of the American Geriatrics Society* 46(2):251-252, 1998.
- Bernabei, R., Gambassi, G., Lapane, K., et al.: Management of Pain in Elderly Patients with Cancer. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology. *Journal of the American Medical Association* 279(23):1877-1882, 1998.
- Davis, M.A.: On Nursing Home Quality: A Review and Analysis. *Medical Care Review* 48(2):129-166, 1991.
- Frederiksen, K., Tariot, P., and De Jonghe, E.: Minimum Data Set Plus (MDS+) Scores Compared with Scores from Five Rating Scales. *Journal of the American Geriatric Society* 44(3):305-309, March 1996.
- Fries, B.E., Simon, S.E., Morris, J.N., et al.: Pain in U.S. Nursing Homes: Validating a Pain Scale for the Minimum Data Set. *The Gerontologist* 41(2):173-179, 2001.
- Gambassi, G., Landi, F., Peng, L., et al.: Validity of Diagnostic and Drug Data in Standardized Nursing Home Resident Assessments: Potential for Geriatric Pharmacoepidemiology. *Medical Care* 36(2):167-179, February 1998.
- Hartmaier, S. L., Sloane, P. D., Guess, H. A., and Koch, G. G.: The MDS Cognition Scale: a Valid Instrument for Identifying and Staging Nursing Home Residents with Dementia Using the Minimum Data Set. *Journal of the American Geriatrics Society* 42(11):1173-1179, 1994.
- Harrington, C., Carrillo, H., Thollaug, S.C., Summers, P.R.: *Nursing Facilities, Staffing, Residents, and Facility Deficiencies, 1991 Through 1997*. Report produced by the Department of Social and Behavioral Sciences, University of California at San Francisco. 1999a.
- Harrington, C., Zimmerman, D., Karon, S., et al.: *Nursing Home Staffing and its Relationship to Deficiencies*. Report produced by the Department of Social & Behavioral Sciences, University of California at San Francisco, and the Center for Health Systems Research and Analysis, University of Wisconsin-Madison. 1999b.
- Hawes, C., Morris, J.N., Phillips, C.D., et al.: Reliability Estimates for the Minimum Data Set for Nursing Home Resident Assessment and Care Screening. *Gerontologist* 35(2):172-178; April 1995.
- Kane, R.A., Kane, R.L., Giles, K., et al.: Measures, Indicators, and Improvement of Quality of Life in Nursing Homes: First Findings from Wave 1 Data Collection. Report prepared for the Centers for Medicare & Medicaid Services, October 2000.
- Kramer, A.: Manual of Operations for Staged Quality of Care Survey. University of Colorado Health Sciences Center, 1999.
- Landi, F., Gambassi, G., Lapane, K. L., et al.: Comorbidity and Drug Use in Cognitively Impaired Elderly Living in Long-Term Care. *Dementia and Geriatric Cognitive Disorders* 9(6):347-356, November-December 31, 1998.
- Landi, F., Gambassi, G., Lapane, K. L., et al.: Impact of the Type and Severity of Dementia on Hospitalization and Survival of the Elderly. *Dementia and Geriatric Cognitive Disorders* 10:121-129, 1999.
- Morris, J.N., Fries, B.E., Mehr, D.R, et al.: MDS Cognitive Performance Scale. *Journal of Gerontology* 49(4):M174-182, July 1994.
- Morris, J.N., Hawes, C., Fries, B.E., et al.: Designing the National Resident Assessment Instrument for Nursing Homes. *Gerontologist* 30(3):293-307, June 1990.
- Morris, J.N., Nonemaker, S., Murphy, K., et al.: A Commitment to Change: Revision of CMS's RAI. *Journal of the American Geriatrics Society* 45(8):1011-1016, August 1997.
- Nyman, J.A.: Improving the Quality of Nursing Home Outcomes. Are Adequacy- or Incentive-Oriented Policies More Effective? *Medical Care* 26(12):1158-1171, December 1988.
- Zimmerman, D., Karon, S., and Swearingen, J.: Report on *The Quality Component of the Demonstration*, in DRAFT: Final Report to Health Care Financing Administration NHCMQ Demonstration Final Report. National Technical Information Service. Springfield, VA. 1999.
- Zimmerman, D., and Karon, S.: *Measuring the Quality of Nursing Home Care: Risk Adjustment, Validation, and other "Trivial" Issues*. Presentation to the National Case Mix Reimbursement and Quality Assurance Conference. September 22, 1997.
- Zimmerman, D., Karon, S., Arling, G., et al.: Development and Testing of Nursing Home QIs. *Health Care Financing Review* 19(4):107-127, 1995.

Zinn, J.S.: Market Competition and the Quality of Nursing Home Care. *The Journal of Health Politics Policy Law* 19(3):555-582, Fall 1994.

Zinn, J.S., Aaronson, W.E., and Rosko, M.D.: The Use of Standardized QIs as Quality Improvement Tools: An Application in Pennsylvania Nursing Homes. *American Journal of Medical Quality* 8(2):72-78, Summer 1993a.

Zinn, J.S., Aaronson, W.E., and Rosko, M.D.: Variations in the Outcomes of Care Provided in Pennsylvania Nursing Homes. Facility and Environmental Correlates. *Medical Care* 31(6):475-487, June 1993b.

---

Reprint Requests: Katherine Berg, Ph.D., McGill University School of Physical and Occupational Therapy, 3654 Sir William Osler Promenade, Montreal, Quebec, Canada H3G 1Y5. E-mail: Katherine.Berg@mcgill.ca